

Relacionamento entre a oferta e a demanda em software para o agronegócio

Stanley Robson de Medeiros Oliveira

Maria Fernanda Moura

Anderson Rodrigo dos Santos

Claudia Juliana Poker Moretti

Thiago Romano dos Santos

4.1 Introdução

Após a apresentação, no capítulo anterior, dos dados e das análises sobre o mercado ofertante de software agropecuário no Brasil, com o recorte das 162 empresas privadas participantes da pesquisa e das 19 unidades da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) que também atuam nesse segmento de mercado, este capítulo avança no estudo abordando o relacionamento entre a oferta e a demanda em software para o agronegócio.

O objetivo do capítulo é realizar uma comparação entre o mercado ofertante de software agrícola e as demandas em software rural identificadas junto a dois agentes do agronegócio: 230 cooperativas agrícolas e 132 órgãos da Assistência Técnica e Extensão Rural (ATER)¹.

O relacionamento entre a oferta e a demanda não é exaustivo. O que se pretende é que essa investigação possa estimular a realização de estudos que contemplem temas tais como:

- Panorama do mercado brasileiro ofertante de software para o agronegócio, destacando as empresas privadas desenvolvedoras e/ou distribuidoras e os seus produtos.
- Detalhamento das empresas demandantes de soluções em tecnologia da informação e comunicação (TIC) no universo agropecuário, incluindo suas necessidades na forma de produtos e serviços.
- Rodada de negócios com o objetivo de ativar as redes de serviços de TIC e articular a oferta de produtos de software prestadores de serviços de cada localidade e as empresas demandantes de soluções.
- Identificação dos nichos de mercado e necessidade de realinhamento dos projetos das empresas desenvolvedoras de software rural para os próximos anos.

¹ Em decorrência da dimensão geográfica do Brasil e do orçamento limitado do projeto, não foi viável estender o levantamento de demandas para mais agentes.

Por fim, espera-se que essa pesquisa possa contribuir de alguma forma para a realização de futuros estudos sobre os cenários de adoção de tecnologia da informação pelo setor rural.

O capítulo está estruturado em 5 seções incluindo esta introdução. A seção a seguir aborda o marco teórico sobre segmentação de mercado em geral, para, num segundo momento, efetuar uma aproximação do mercado de software para o agronegócio, segmentando-o em demandantes e ofertantes. A seção seguinte trata da metodologia utilizada para segmentação dos demandantes e ofertantes, a Cross Industry Standard Process for Data Mining (CRISP-DM). Na seção subsequente, são analisados os perfis, ou seja, segmentos de demandas em software agropecuário entre as cooperativas rurais e os órgãos de assistência técnica e extensão rural, e apresentados os perfis dos ofertantes de software agropecuário, bem como discutidos os possíveis nichos de mercado que esses segmentos não estão cobrindo. Por último, são discurridas as considerações finais.

4.2 Segmentação de mercado à luz da literatura

Avanços tecnológicos recentes têm provocado profundas mudanças e adaptações em empresas, produtos e serviços. O ritmo acelerado dessas mudanças está impactando segmentos agrícolas, até os mais conservadores, que pouco consideraram a necessidade ou importância do suporte tecnológico (COSTA; OLIVEIRA, 1998).

O cenário atual revela a crescente necessidade por informação de qualidade como subsídio essencial aos processos de gestão. À proporção em que o agronegócio avança, percebe-se a necessidade de se implementar medidas eficientes para que sejam superados os obstáculos identificados ao seu bom desempenho (BRANDÃO; MEDEIROS, 1998). Essas medidas, relacionadas à gestão da informação e à tomada de decisão para o agronegócio, demandam uma infraestrutura de tecnologia de comunicação eficiente, uma vez que o cenário na geração de produtos e serviços mais competitivos, em atendimentos às necessidades de mercado crescente e de forte concorrência, requerem a adoção de novas tecnologias para automação de seus processos (COUTINHO; FERRAZ, 1994).

Uma das características marcantes da tecnologia da informação é sua abrangência. Suas aplicações cobrem todos os setores da economia mesmo que de forma e intensidade diferenciadas em cada setor ou até entre atividades dentro de um mesmo setor. Se suas aplicações apareceram inicialmente de maneira mais rápida e generalizada nos setores terciários e secundários, não tardaram a se expandir para o setor primário. Na agricultura, a tecnologia da informação pode ser compreendida como mais uma etapa no processo de modernização (ARRAES; LYRA FILHO, 1995).

Embora a tecnologia da informação englobe diversas formas de aplicação, a tecnologia de gerenciamento de informações por meio de computadores e aplicativos específicos são as que, aparentemente, têm sido mais adotadas. Até porque, por alguns anos, sua oferta foi quase que exclusivamente em relação aos outros tipos de aplicação (automação, telemática, robótica). Contudo, um dos desafios presentes no cenário atual é conhecer como tem se dado o processo de difusão da tecnologia da informação na agricultura brasileira por meio da análise da oferta de programas computacionais comerciais para aplicações em atividades rurais, desenvolvidos por empresas do mercado brasileiro de software agropecuário.

Um pré-requisito relevante que surge ao se decidir fortalecer a difusão da tecnologia da informação na agricultura brasileira é a disponibilidade de informações sistematizadas sobre

ofertantes e demandantes do mercado de software agropecuário. São elas que permitem o conhecimento detalhado da realidade das organizações que atuam no domínio agropecuário, com possibilidades de atualização permanente das informações, de modo que se construam as bases para a identificação de nichos de mercado e necessidade de realinhamento dos projetos das empresas desenvolvedoras de software rural para os próximos anos.

Uma alternativa para abordar esse problema seria a segmentação dos demandantes e ofertantes do mercado de software agropecuário brasileiro, com o objetivo de se conhecer melhor suas necessidades na forma de produtos e serviços em tecnologia da informação e, em seguida, tentar articular as empresas desenvolvedoras de software que possam suprir essa demanda. Nesse contexto, entende-se que mercado é um determinado local onde os demandantes e os ofertantes se reúnem para realizar as trocas de bens ou serviços, no âmbito da tecnologia da informação.

Segmentar um mercado significa escolher um grupo de consumidores, com necessidades homogêneas, para o qual uma empresa poderá fazer uma oferta mercadológica. O processo de segmentação requer que sejam identificados os fatores que afetam as decisões de compras dos consumidores (KOTLER; ARMSTRONG, 1999). Para satisfazer os requisitos para a segmentação, o seguimento deve ser identificável, mensurável, acessível, rentável e estável.

Na segmentação, a meta analítica é medir o comportamento do consumidor e colocar cada instituição num grupo (segmento) que minimize a variação de comportamento entre os membros do segmento e maximize a variação entre os segmentos (WEINSTEIN, 1995). As variáveis mais frequentemente utilizadas como base para segmentação do mercado de consumo são classificadas por Kotler (1998) como: geográficas, demográficas, psicográficas e comportamentais.

Segundo Weinstein (1995), a abordagem de segmentação proposta por Kotler (1998) possui vários benefícios: (1) Aplicação abrangente – é um tipo de segmentação apropriado a mercados de bens de consumo, industriais e internacionais; (2) Bases causais – diferentemente de outras bases de segmentação, os segmentos por benefícios são baseados em fatores de causa e efeito muito mais do que em fatores descritivos. Como os segmentos comportamentais identificam por que os consumidores compram, existe uma relação direta entre motivações e padrões de compra; e (3) Flexibilidade – é um método de grande adaptabilidade, podendo ser utilizado em colaboração com muitas outras bases de segmentação intimamente correlacionadas.

Para Kotler (1999), os mercados consistem em compradores que se diferem de várias formas quanto aos seus desejos, recursos, localidades, atitudes e práticas de compra. Portanto, a segmentação de mercado é o processo de projetar ou caracterizar um produto ou serviço que exercerá uma atração especialmente forte para alguma subparte identificável no mercado total.

Ainda com relação à segmentação de mercado, Rocha e Christensen (1999) enfatizam que o processo se baseia em desdobramentos do lado da demanda e representa um ajuste racional e mais preciso do produto e do esforço de marketing às exigências do consumidor ou usuário.

De uma forma geral, pode-se dizer que a segmentação de mercado consiste na tarefa de identificar e classificar grupos distintos de ofertantes e de demandantes que podem exigir produtos. Em particular, o escopo de produtos considerado nessa avaliação contempla software agropecuário e serviços de tecnologia da informação.

No processo de segmentação em estudo, os compradores representam as instituições que demandam software agropecuário, formadas por cooperativas agrícolas e as instituições de assistência técnica e extensão rural, ao passo que os ofertantes são constituídos de empresas que produzem e/ou comercializam software agropecuário.

4.3 Metodologia

Com o objetivo de facilitar a segmentação dos demandantes e ofertantes de software agropecuário, foi utilizada a metodologia CRISP-DM (CRoss Industry Standard Process for Data Mining) (CHAPMAN et al., 2000). O processo de análise de dados CRISP-DM, conforme Figura 4.1, consiste em seis fases: compreensão do domínio, entendimento dos dados, preparação dos dados, modelagem, avaliação e distribuição.

O ciclo externo, na Figura 4.1, simboliza a natureza cíclica da análise exploratória de dados. O processo não termina uma vez que uma solução é encontrada. As lições aprendidas durante o processo podem gerar novos questionamentos, geralmente mais pertinentes ao assunto. Processos subsequentes se beneficiarão das experiências adquiridas nos processos anteriores.

Fase 1 – Compreensão do domínio: O processo inicia-se com a compreensão do domínio, buscando obter conhecimento sobre seus objetivos e requisitos para defini-lo e convertê-lo em um problema de análise de dados, traçando-se, então, um plano preliminar designado para alcançar seus objetivos.

Conforme discutido na introdução, este trabalho consiste na realização de um estudo comparativo entre o mercado ofertante de software agrícola e as demandas em tecnologia da informação identificadas junto às cooperativas agrícolas e às instituições de assistência técnica e extensão rural.

Fase 2 – Entendimento dos dados: A fase dois inicia-se com o levantamento dos dados para avaliar os possíveis problemas de qualidade, identificando subconjuntos relevantes. Em segui-

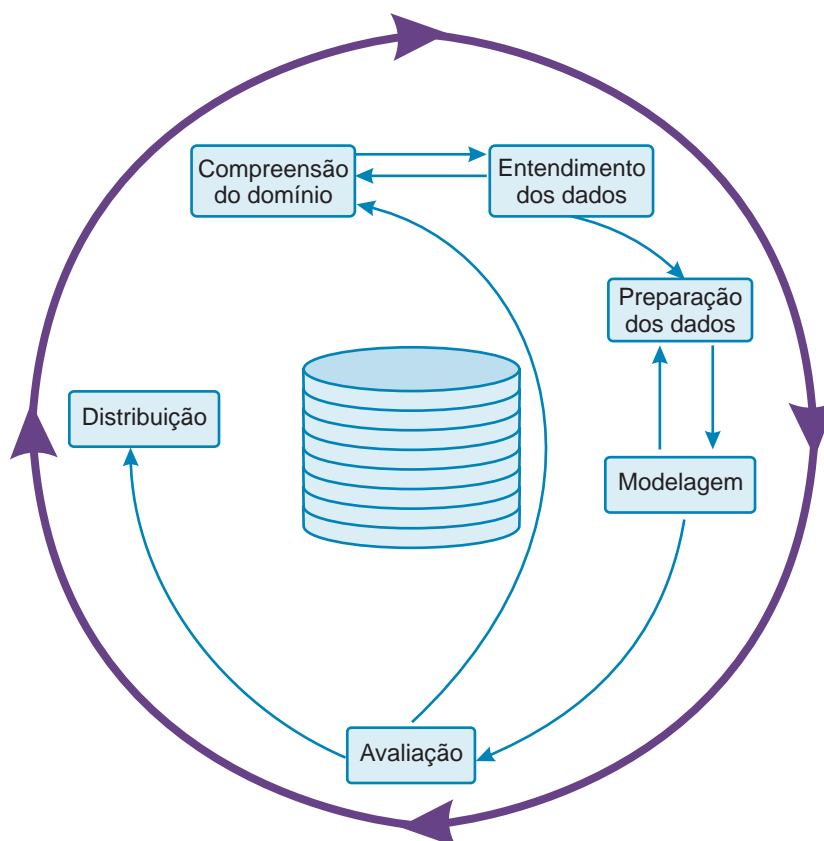


Figura 4.1. Fases do processo CRISP-DM.

Fonte: Chapman et al. (2000), adaptada pelo autor.

da, alguns procedimentos e atividades são aplicados a esses dados, visando identificar problemas de qualidade e detectar subconjuntos interessantes para formalizar hipóteses.

O levantamento das empresas que desenvolvem/comercializam programas para o setor agropecuário foi realizado de 2008 a 2010, pela Embrapa Informática Agropecuária, em parceria com diversas instituições. Tanto o levantamento da oferta como o da demanda de software para o agronegócio foram realizados em duas etapas. A primeira foi a elaboração de um formulário semiestruturado *online* para preenchimento dos dados pelas instituições participantes da pesquisa. Para orientar o preenchimento dos dados, foi disponibilizado um mecanismo de ajuda com instruções sobre o preenchimento de cada campo do questionário. A segunda etapa refere-se à organização das informações em um banco de dados para armazenar e gerar relatórios sobre as informações levantadas.

Como apresentado no capítulo anterior, entre as ofertantes foram identificadas 162 empresas desenvolvedoras para 402 produtos de software, conforme Tabela 4.1; na qual pode-se encontrar a frequência absoluta e relativa por unidade da federação. Além desse número de empresas, acrescentam-se 19 Centros de Pesquisa da Embrapa, responsáveis por 68 produtos de software, relacionados na Tabela 4.2 por unidade da federação na qual esses centros se localizam.

Com relação aos demandantes, 230² cooperativas agrícolas manifestaram interesse em participar do levantamento juntamente com 132 órgãos estaduais de assistência técnica e extensão

Tabela 4.1. Distribuição das empresas privadas, desenvolvedoras de software participantes do projeto e suas respectivas unidades da federação (em 2010).

Local	Quantidade	%
SP	53	32,7
MG	34	21,0
PR	23	14,2
RS	13	8,0
SC	9	5,6
MT	8	4,9
PE	4	2,5
ES	3	1,9
GO	3	1,9
MS	3	1,9
RJ	3	1,9
DF	2	1,2
SE	1	0,6
BA	1	0,6
CE	1	0,6
PA	1	0,6
Total	162	100,00

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

Tabela 4.2. Centros de Pesquisa da Embrapa desenvolvedores de software, segundo unidades da federação (em 2010).

Unidades da Embrapa	UF
Embrapa Acre	AC
Embrapa Agroindústria Tropical	CE
Embrapa Agropecuária Oeste	MS
Embrapa Amazônia Oriental	PA
Embrapa Cerrados	DF
Embrapa Florestas	PR
Embrapa Gado de Corte	MS
Embrapa Gado de Leite	MG
Embrapa Hortaliças	DF
Embrapa Informática Agropecuária	SP
Embrapa Instrumentação Agropecuária	SP
Embrapa Meio Ambiente	SP
Embrapa Meio-Norte	PI
Embrapa Pecuária Sudeste	SP
Embrapa Soja	PR
Embrapa Solos	RJ
Embrapa Suínos e Aves	SC
Embrapa Tabuleiros Costeiros	SE
Embrapa Transferência de Tecnologia	DF
Total	19

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

2 Segundo informações da Organização das Cooperativas do Brasil (OCB), em dezembro/2010, havia 1519 cooperativas rurais associadas à OCB. As 230 que participantes da pesquisa representam 15,14%.

rural. A Tabela 4.3 apresenta a distribuição geográfica das cooperativas por estado e região, enquanto a Tabela 4.4 mostra a distribuição geográfica dos órgãos de extensão rural. Nas duas tabelas, encontram-se as frequências absolutas e relativas da distribuição dos demandantes em cada unidade da federação e região. Como esperado, há uma concentração geográfica de cooperativas nas regiões Sudeste e Sul; bem como, as empresas de extensão rural estarem concentradas no Sudeste. Porém, cabe ressaltar que houve um maior número de órgãos de assistência técnica e extensão rural participantes dessa pesquisa localizados no Estado de São Paulo, o que pode ter ocasionado um viés dos resultados.

Fase 3 – Preparação dos dados: O terceiro passo trata da preparação dos dados, em que todas as atividades são voltadas para a construção do banco de dados no formato adequado para o

Tabela 4.3. Distribuição das 230 Cooperativas rurais em 26 estados da federação, agrupadas por região (2010).

Cooperativas rurais					
AL	4	1,74%	AM	2	0,87%
AP	2	0,87%	BA	16	6,96%
CE	3	1,30%	DF	3	1,30%
ES	8	3,48%	GO	10	4,35%
MA	1	0,43%	MG	43	18,70%
MS	9	3,91%	MT	3	1,30%
PA	2	0,87%	PB	7	3,04%
PE	4	1,74%	PI	2	0,87%
PR	15	6,52%	RJ	6	2,61%
RN	4	1,74%	RO	3	1,30%
RR	3	1,3%	RS	35	15,22%
SC	18	7,83%	SE	1	0,43%
SP	24	10,43%	TO	2	0,87%
Cooperativas por região					
Norte	14	6,09%	Nordeste	42	18,26%
Sul	68	29,57%	Sudeste	81	35,22%
Centro-Oeste	25	10,87%			

Fonte: Banco de dados da demanda em software agropecuário, do projeto SW Agro.

Tabela 4.4. Distribuição dos órgãos estaduais de assistência técnica e extensão rural segundo região por unidades da federação (2010).

Órgãos estaduais de Assistência Técnica e Extensão Rural identificados					
AP	1	0,76%	BA	1	0,76%
DF	1	0,76%	PR	1	0,76%
RJ	1	0,76%	RO	5	3,79%
SC	1	0,76%	SP	120	90,91%
TO	1	0,76%			
Órgãos estaduais de Assistência Técnica e Extensão Rural por região					
Norte	7	5,30%	Nordeste	1	0,76%
Sul	2	1,52%	Sudeste	121	91,67%
Centro-Oeste	1	0,76%			

Fonte: Banco de dados da demanda em software agropecuário, do projeto SW Agro.

processo de extração do conhecimento. As tarefas incluem limpeza, transformação, integração e formatação dos dados. É nessa etapa que possíveis dados inconsistentes são eliminados.

Os dados dos ofertantes e demandantes de software agropecuário, coletados por meio dos questionários *online*, foram armazenados em diversas tabelas, integrados em um banco de dados relacional³. Três conjuntos de dados (um para cooperativas agrícolas, outro para Órgãos Estaduais de Assistência Técnica e Extensão Rural, e o último para ofertantes de software agropecuário) foram selecionados, por especialistas, para serem analisados na fase posterior. Os atributos de cada conjunto de dados foram selecionados de acordo com sua contribuição para a segmentação (clusterização ou agrupamento de dados) de ofertantes e demandantes.

Os atributos das cooperativas amostradas segundo a demanda por software agropecuários são mencionados na Seção 4.4.1, enquanto os atributos dos Órgãos Estaduais de Assistência Técnica e Extensão Rural, amostrados segundo a demanda por software agropecuário, são relatados na Seção 4.4.2. Por fim, na Seção 4.4.3, encontra-se a lista dos atributos das empresas ofertantes de software agropecuário.

Os dados analisados foram obtidos dos bancos de dados da oferta e da demanda em software agropecuário, armazenados na Embrapa Informática Agropecuária, coletados durante a execução do projeto Estudo do Mercado Brasileiro de Software Agropecuário (SW Agro), com acesso restrito à equipe técnica.

Convém destacar que nos conjuntos de dados selecionados para ofertantes e demandantes não foram observados dados faltantes (*missing values*).

Fase 4 – Modelagem: Nessa etapa, técnicas de segmentação de dados foram selecionadas e aplicadas aos conjuntos de dados selecionados na fase anterior. Geralmente, existem várias técnicas para o mesmo tipo de problema de segmentação de dados. Algumas técnicas possuem requisitos específicos na forma dos dados. Consequentemente, voltar para a etapa de preparação de dados é frequentemente necessário.

O processo de agrupamento de objetos em classes similares é denominado clusterização. Esse procedimento é também conhecido como segmentação de dados, pois particiona grandes conjuntos de dados de acordo com a similaridade entre subconjuntos. É geralmente utilizado como técnica inicial de exploração dos dados, com o objetivo de agrupá-los em um número finito de conjuntos semelhantes, conhecidos como *clusters*.

Nesse procedimento, os objetos são agrupados baseados no princípio da maximização da variabilidade dos dados intergrupos e da minimização intragrupos (HAN; KAMBER, 2006). Sendo assim, objetos pertencentes ao mesmo *cluster* são altamente similares entre si, e amplamente dissimilares em relação a qualquer outro objeto de outros *clusters*. Nesse processo, não é necessária a identificação dos agrupamentos desejados ou dos atributos que devem ser usados na criação dos *clusters*.

De acordo com Jain et al. (1999), o processo de clusterização envolve basicamente:

- Representação dos padrões (pode incluir extração ou seleção de características): refere-se à descrição dos atributos, ou características observadas (número, tipo e modo). A seleção de atributos refere-se ao processo de identificação do melhor subconjunto a ser usado na clusterização. A extração de atributos trata do uso de uma ou mais transformações nos atributos de entrada para salientar novas características.
- Definição de uma medida de similaridade apropriada para o domínio de aplicação: medidas de distância são utilizadas para compor uma medida de similaridade dentre os elementos

³ Um Banco de Dados Relacional é um conceito abstrato que define maneiras de armazenar, manipular e recuperar dados estruturados unicamente na forma de tabelas, construindo um banco de dados.

dos *clusters* obtidos. Entre as mais comuns, pode-se citar a distância Euclidiana, a distância Manhattan, etc.

- Clusterização ou agrupamento: O processo de clusterização pode ser feito pela partição dos dados em grupos, ou pela medida do grau de pertinência de um dado em relação a um conjunto formado.
- Abstração dos dados: é o processo de extração de uma representação simples e compacta do conjunto de dados, como por exemplo, a definição dos centroides, elementos que representam o centro de um grupo, para cada *cluster* obtido.
- Apresentação dos resultados: visualização dos *clusters* e compreensão de suas inter-relações.

A Figura 4.2 representa a sequência dos três primeiros passos, incluindo um *feedback*, em que os dados de saída podem realimentar processos anteriores.

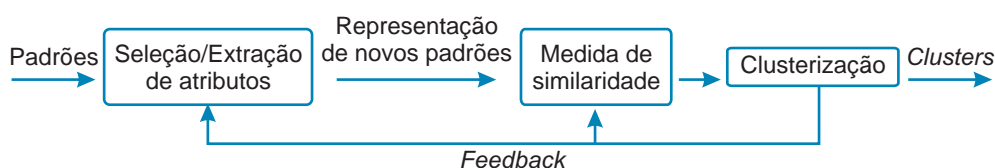


Figura 4.2. Estágios do processo de clusterização.

Fonte: Jain et al. (1999), adaptada pelo autor.

A definição do número ideal de *clusters* é uma tarefa subjetiva, ou empírica, que depende das necessidades do usuário ou de sua experiência e perspicácia.

Método para segmentação de dados

Existem diversos métodos para segmentação descritos na literatura (HAN; KAMBER, 2006). Os principais métodos são: particionamento, hierárquicos e baseados em densidade de probabilidade. A escolha de um método depende do tipo de dado a ser analisado, assim como do propósito e da aplicação dessa análise. Em particular, neste estudo foi selecionado um método de particionamento.

Método de particionamento

Quando há um conjunto de dados contendo n objetos, o método de particionamento divide o conjunto em k partes, correspondentes aos *clusters*, sendo $k \leq n$. Os grupos obtidos devem conter no mínimo um objeto, e cada objeto deve pertencer a apenas um *cluster*. Dado o valor de k , os objetos são alocados aos *clusters* de acordo com a medida de similaridade adotada, buscando a maximização da distância entre *clusters* e a minimização da distância entre objetos do mesmo grupo. De acordo com Han e Kamber (2006), existem inúmeras possibilidades na determinação do centroide (elemento que representará o centro do *cluster*) e que, consequentemente, servirá como base para o cálculo da medida de distância, tendo-se, como exemplo, o algoritmo k-means.

O k-means (também conhecido por k-médias) é uma técnica na qual os dados são agrupados de acordo com uma medida de distância especificada para o algoritmo, conforme a representatividade das características dos dados, se qualitativa ou quantitativa. Primeiramente, o algoritmo seleciona objetos aleatoriamente, de acordo com o número de *clusters* (k) definido pelo usuário. Esses objetos são representantes dos centroides. Em seguida, k *clusters* são criados por meio da associação de cada objeto ao centroide mais próximo. Esse passo refere-se ao cálculo da distância entre os elementos. Posteriormente, calcula-se o valor médio da distância de cada objeto pertencente ao *cluster*; com isso, novos centroides são selecionados e, conse-

quentemente, uma nova distribuição é realizada. Esse processo se repete até a obtenção do melhor conjunto, visando à minimização da distância entre os objetos do mesmo conjunto e a maximização em relação aos outros grupos formados (AMO, 2004; REZENDE et al., 2005). A Figura 4.3 ilustra o processo de segmentação de dados usando o algoritmo k-means. Nesta figura, os pontos vermelhos representam os centroides, que são recalculados após cada iteração do algoritmo.

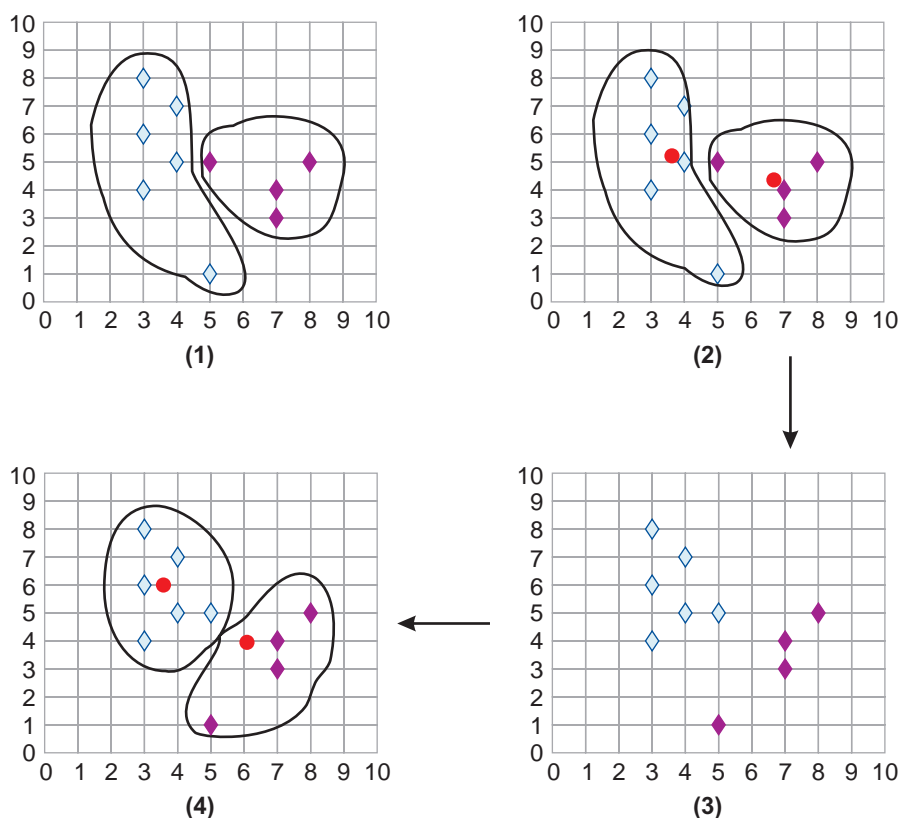


Figura 4.3. Exemplo do processo de segmentação de dados usando o algoritmo k-means.

Fonte: Han e Kamber (2006), adaptada pelo autor.

Basicamente se realiza o cálculo do representante (centroide) do agrupamento como a média dos atributos reais e a moda⁴ dos atributos categóricos. O cálculo da distância entre dois objetos, ou entre um objeto e um centroide, é realizado como a distância euclidiana dos atributos contínuos somada à distância de Hamming dos atributos categóricos.

O algoritmo k-means possui um bom desempenho quando os clusters são densos e bem separados uns dos outros. Esse método é relativamente escalável no processamento de grandes bases de dados, uma vez que o número de clusters gerados é bem menor que o número total de objetos, assim como é menor que o número de iterações necessárias para formação dos clusters.

Um ponto fraco desse algoritmo é a necessidade de especificação do número de clusters com antecedência, o que pode ser uma desvantagem, uma vez que o número escolhido pode não ser a melhor forma de divisão dos dados. Outro, é que esse método é sensível a ruídos e *outliers*, já

⁴ Em estatística descritiva, a moda é o valor que detém o maior número de observações, ou seja, o valor ou valores mais frequentes. A moda não é necessariamente única, ao contrário da média ou da mediana.

que todos os objetos pertencentes a um conjunto de dados são agrupados e, assim, um número pequeno de tais dados (*outliers*) pode influenciar o valor do centroide.

O problema da definição do número de grupos

Uma das grandes dificuldades nos problemas de agrupamento, como no método clássico k-means, certamente é a exigência do valor de k, o número de segmentos ou grupos em que a massa de dados será dividida, como parâmetro inicial. Para uma aplicação real, em grandes bases de dados, essa informação normalmente é desconhecida, podendo ser necessária uma análise prévia de um especialista para tentar identificar a quantidade de grupos em que a coleção poderia se dividir de forma satisfatória, o que por si só já representa uma grande carga de trabalho que, até mesmo, diante de uma base muito grande, poderia tornar a análise inviável ou desinteressante.

Existem vários critérios para a determinação do número de grupos e quase todos funcionam da seguinte maneira: realizar o agrupamento dos dados considerando 2 grupos e calcular o valor de uma função proposta que tenha o número de grupos como um de seus parâmetros, realizar o agrupamento dos dados considerando 3 grupos e calcular o valor da mesma função, repetir esse procedimento até atingir um número máximo de grupos estabelecidos. O agrupamento que ocasionar o valor máximo (ou, em alguns casos, mínimo) da função, deve ser considerado como o melhor agrupamento possível para a base de dados.

Há diversos índices semi-empíricos que se pode usar, tais como, Calinski e Harabasz, Critério Condorcet, Cubic Clustering Criterion e PBM, para ver qual o melhor “k”, número de grupos. Para esse tipo de classificação não supervisionada, há uma vasta literatura, já que ele tem sido tema de pesquisa ininterrupta (ANDRADE, 2004; MACHADO FILHO, 2002; MORAES, 2004; PUNTAR, 2003).

Nesse estudo, adotou-se o algoritmo Expectation-Maximization (EM) para se estimar o número de grupos (*clusters*). Esse algoritmo estende o paradigma usado no k-means, associando uma distribuição de probabilidade a cada instância, que indica a probabilidade dessa instância pertencer a uma única população e, conseqüentemente, encontra o número estimado de populações na amostra. Em geral, o algoritmo EM é aplicado em situações onde se deseja estimar um conjunto de parâmetros que descreve uma distribuição de probabilidade (DEMPSTER et al., 1977), isto é, ele estima a média amostral e sua variância. O algoritmo calcula os estimadores de máxima verossimilhança para problemas onde existem dados incompletos entre os dados observados ou quando os estados das variáveis não foram observados, dados não observados.

Cada iteração do algoritmo EM envolve dois passos que são: expectation (passo E) e maximization (passo M). O Passo E desse algoritmo consiste em obter uma estimativa da esperança (expectation) dos dados faltosos para completar a amostra de dados incompleta, partindo-se de um valor inicial empírico. No Passo M, com os dados completados, realiza-se aprendizagem das probabilidades, baseado nas frequências dos estados das variáveis na amostra, de modo a calcular a função de maximização da esperança encontrada. Os Passos E e M fazem parte de um processo iterativo, em que as novas probabilidades, calculadas na fase M, serão utilizadas para realizar a inferência na fase E.

Fase 5 – Avaliação: A penúltima fase corresponde à avaliação do processo, em que se deve construir um modelo que pareça de alta qualidade, numa perspectiva de análise de dados. Antes de prosseguir é importante avaliar detalhadamente o modelo, e rever os passos executados na sua construção, para certificar-se de que ele alcançará os objetivos propostos. Deve-se, ainda, determinar se não há algum ponto importante, referente ao assunto, que não tenha sido suficientemente abordado. No fim dessa fase, uma decisão sobre o uso dos resultados deve ser tomada.

O procedimento para segmentação dos ofertantes e demandantes foi realizado em duas etapas. Na primeira, o algoritmo EM foi aplicado a cada conjunto de dados para se obter o número adequado de *clusters* (grupos) a serem gerados pelo modelo, ou seja, para estimar o número de populações na amostra. Em seguida, o algoritmo k-means foi aplicado para determinar os *clusters* e seus respectivos centroides, considerando-se k como o número de populações estimadas. Embora o EM sozinho já calcule os grupos, o uso do k-means, como implementado no ambiente WEKA (WITTEN; FRANK, 2005), traz o ganho de fornecer descritores para cada grupo, facilitando a interpretação deles; além disso, ele trata de uma maneira mais precisa a questão de misturas de atributos quantitativos (reais ou inteiros) e qualitativos (binários, categóricos ou nominais).

Fase 6 – Distribuição: Na última fase, tem-se um modelo construído. No entanto, isso geralmente não representa o fim do projeto. Mesmo se o propósito do modelo for aumentar o conhecimento dos dados, o conhecimento adquirido necessitará ser organizado e apresentado de uma maneira que o usuário possa utilizá-lo. Isso geralmente envolve procedimentos de validação com especialistas nos processos de tomada de decisão. Dependendo das exigências, a aplicação do modelo pode ser tão simples quanto a geração de um relatório ou tão complexa quanto executar processos de análise de dados, repetidamente. Em muitos casos, será o usuário não especialista (não o analista dos dados) que realizará os passos para execução do modelo, por isso é necessário que ele compreenda que medidas deverão ser tomadas a fim de empregar efetivamente os modelos criados.

As fases 5 e 6 são abordadas com mais detalhes no decorrer da seção seguinte (Resultados obtidos).

4.4 Resultados obtidos

Essa seção está dividida em três partes. Primeiro são analisados os possíveis perfis, ou seja, segmentos de demandas em software agropecuário entre as cooperativas rurais que responderam aos questionários. Na sequência, foram analisados os perfis dos órgãos de assistência técnica e extensão rural e os resultados são discutidos juntamente com aqueles obtidos das cooperativas. Finalmente são apresentados os perfis dos ofertantes de software agropecuário e discutidos os possíveis nichos de mercado que esses segmentos não estão cobrindo.

4.4.1 Perfis das cooperativas rurais amostradas segundo a demanda por software agropecuário

Nessa seção mostra-se como foram agrupadas as cooperativas rurais amostradas, segundo o seu perfil de uso e demanda de software agropecuário. Procurou-se verificar quais eram as características dos produtos de software em uso pelas cooperativas e quais as principais dificuldades que têm sido encontradas. A seguir, o interesse concentrou-se nos investimentos pretendidos por elas e, finalmente, em quais seriam os focos de demanda, por área de aplicação. Em um primeiro momento, essa análise não separou muito bem os perfis das cooperativas, então se acrescentaram os dados sobre manejo animal e cultivo vegetal de foco das cooperativas.

Então, embora se tenham apenas 230 cooperativas amostradas, cada uma dessas cooperativas poderia apresentar mais de um perfil, dependendo de como se combinem esses diversos

fatores. Para avaliar os perfis, o conjunto final de características consideradas foi: a UF da cooperativa possuir ou não computador, ter ou não acesso à internet, disponibilizar ou não o acesso à internet, utilizar ou não software agropecuário, problemas encontrados no uso de software agropecuário, pretender ou não fazer investimentos em tecnologia da informação, os motivos de realizar investimentos, o tipo de área de aplicação de software agropecuário demandado pela cooperativa (administração ou controle de processos), a área de aplicação do software, o tipo de manejo animal e cultivo predominantes entre os cooperados. Logo, para a análise de perfil, as possíveis combinações entre as cooperativas amostradas e cada conjunto de características observadas chegaram a 15.711 indivíduos obtidos (elementos a serem agrupados).

Aplicando-se o método EM para verificar quantos perfis ou segmentos teríamos nessa amostra, o resultado foi 18 perfis; e, a seguir, aplicando-se o método k-means, com distância de Manhattan (pois a maioria dos dados é qualitativo, valores binários: sim ou não), e conhecido o k (18 perfis), temos as modas de cada característica para cada perfil na Tabela 4.5. Como algumas características foram comuns aos 18 perfis, elas não estão apresentadas na tabela, apenas para simplificar a visualização dela. Essas características foram: possuir computador, acesso à internet, utilizar software e pretender investir; pois nesses casos as modas de todas elas em todos os perfis foram “sim”. Deve-se lembrar de que os valores de modas de cada característica refletem apenas o valor representativo do grupo, isto é, um grupo cuja moda da UF seja RS pode conter uma UF como RN ou AM; a moda reflete o valor predominante da característica no grupo, não necessariamente todos os valores do grupo.

Deve-se notar que, ao fazer uma análise mais subjetiva das características da Tabela 4.5, poder-se-ia dizer que dificuldades como subutilização de funcionalidades do software, complexidade do software e linguagem inadequada, devam ser similares, pois elas levam à necessidade de capacitação de pessoal ou compra de equipamento ou melhoria de infraestrutura, porém de um modo meio combinado. Já outras necessidades voltadas às áreas de aplicação parecem mais relacionadas ao tipo de cultivo ou manejo animal predominante. Por exemplo, o manejo integrado de pragas e agricultura de precisão com a cultura do algodão, a automação agropecuária e o zoneamento com o milho, a rastreabilidade com suínos e a fitossanidade com bovinos de corte.

Observando-se o todo, predominantemente, as cooperativas que apontaram a necessidade de sistemas de software para gerenciamento de equipamentos agrícolas, bem como agricultura de precisão e automação agropecuária, pretendem investir na compra de equipamentos de hardware ou infraestrutura de redes, e que estas se concentram no sul do país. Ou seja, existe a necessidade e há indícios de que haja capital para investir nessas instituições; porém, esse mesmo perfil de cooperativas revela que seus principais problemas no uso de sistemas de software, atualmente, concentram-se em subutilização funcional ou excesso de complexidade do software ou linguagem muito técnica. Isso pode ser um indicativo de necessidade de capacitação, que pode ser fornecida pelos atuais ofertantes de seus softwares. Ainda, há uma maior demanda por sistemas de agricultura de precisão e manejo integrado de pragas quando o cultivo é algodão, com concentração no Paraná. No caso do café, há uma maior demanda por sistemas de controle de irrigação. Sistemas de controle pós-colheita, secagem e armazenamento de grãos aplicam-se mais à soja e ao milho, na amostra de cooperativas analisada.

Tabela 4.5. Modas das características dos 18 perfis de cooperativas agrícolas (2010).

Partição	UF	Problemas encontrados	Motivos investimento	Tipo área aplicação	Área aplicação	Manejo animal	Cultivo vegetal
2673 (17%)	RS	Subutilização das funcionalidades do software	Compra de hardware/computadores/periféricos	Administração/gerenciamento	Gerenciamento/manutenção de maquinários, equipamentos	Bovinos de corte	Milho
1259 (8%)	PR	Subutilização das funcionalidades do software	Compra de hardware/computadores/periféricos	Controle de processos	Agricultura de precisão	Peixes	Algodão
1080 (7%)	PR	Dificuldade ou demora de resposta da assistência técnica quando foi solicitada	Contratar assistência técnica especializada para manutenção de hardware/software	Controle de processos	Manejo integrado de pragas	Bovinos de leite	Algodão
877 (6%)	PR	Falta de documentação técnica sobre o software	Melhoria de infraestrutura interna (por exemplo, implantação de redes)	Administração/gerenciamento	Administração rural	Suínos	Milho
344 (2%)	SC	Sub-utilização das funcionalidades do software	Compra de hardware/computadores/periféricos	Administração/gerenciamento	Gerenciamento de pessoas	Bovinos de leite	Café
1264 (8%)	PR	Complexidade do software	Melhoria de infraestrutura interna (por exemplo, implantação de redes)	Administração/gerenciamento	Gerenciamento/manutenção de maquinários, equipamentos	Bovinos de leite	Café
1215 (8%)	RS	Complexidade do software	Compra de software	Controle de processos	Pós-colheita, processamento e armazenamento de produto	Aves	Soja
654 (4%)	MG	Complexidade do software	Melhoria de infraestrutura interna (por exemplo, implantação de redes)	Controle de processos	Irrigação	Bovinos de leite	Café
336 (2%)	SP	Falta de funcionários treinados	Compra de software	Administração/gerenciamento	Comercialização	Bovinos de corte	Frutas
867 (6%)	PR	Dificuldade ou demora de resposta da assistência técnica quando foi solicitada	Capacitação dos usuários	Controle de processos	Solos (análise química e física)	Bovinos de leite	Trigo
1207 (8%)	RO	Subutilização das funcionalidades do software	Capacitação dos usuários	Controle de processos	Manejo florestal/reflorestamento	Peixes	Sistemas agro florestais
407 (3%)	PR	Subutilização das funcionalidades do software	Melhoria de infraestrutura interna (por exemplo, implantação de redes)	Controle de processos	Zoneamento agrícola	Peixes	Algodão
808 (5%)	PR	Subutilização das funcionalidades do software	Capacitação dos usuários	Controle de processos	Rastreabilidade	Suínos	Soja
392 (2%)	MG	Linguagem/termos muito técnicos	Compra de hardware/computadores/periféricos	Administração/gerenciamento	Administração rural	Bovinos de leite	Feijão

Continua...

Tabela 4.5. Continuação...

Partição	UF	Problemas encontrados	Motivos investimento	Tipo área aplicação	Área aplicação	Manejo animal	Cultivo vegetal
813 (5%)	SC	Falta de funcionários treinados	Compra de hardware/computadores/periféricos	Controle de processos	Rastreabilidade	Suínos	Trigo
545 (3%)	SC	Complexidade do software	Compra de hardware/computadores/periféricos	Administração/gerenciamento	Comercialização	Bovinos de leite	Trigo
626 (4%)	RS	Linguagem/termos muito técnicos	Compra de hardware/computadores/periféricos	Controle de processos	Automação agropecuária (secagem de grãos, beneficiamento de sementes)	Suínos	Milho
344 (2%)	RS	Linguagem/termos muito técnicos	Compra de software	Controle de processos	Fitossanidade	Bovinos de corte	Milho

Fonte: Banco de dados de demandas em software agropecuário, do projeto SW Agro.

4.4.2 Perfis dos órgãos estaduais de assistência técnica e extensão rural amostrados, segundo a demanda por software agropecuário

Nessa seção, mostra-se como foram agrupados os órgãos estaduais de assistência técnica e extensão rural amostrados, segundo o seu perfil de uso e demanda de software agropecuário. Procurou-se verificar quais eram as características dos produtos de software em uso pelas empresas de extensão rural e quais as principais dificuldades encontradas. A seguir, o interesse concentrava-se nas expectativas de investimentos e, finalmente, em quais seriam os focos de demanda por área de aplicação. No caso das empresas de extensão rural, não foram incorporados aos questionários quais eram os tipos de manejo animal e cultivo vegetal de seus clientes. Mesmo assim, para melhor aproximar os perfis, órgãos estaduais de assistência técnica e extensão rural dos perfis identificados entre as cooperativas rurais, geraram-se combinações das empresas de extensão rural amostradas com as características observadas, assim, embora se tenham apenas 132 Aters, cada qual pode apresentar mais de um perfil, dependendo de como se combinem esses diversos fatores. Para avaliar os perfis, o conjunto final de características consideradas foi: a UF da empresa de extensão rural possuir ou não computador, ter ou não acesso à internet, disponibilizar ou não o acesso à internet, utilizar ou não software agropecuário, problemas encontrados no uso de software agropecuário, pretender ou não fazer investimentos em tecnologia da informação, os motivos de realizar investimentos, o tipo de área de aplicação de software agropecuário demandado pela empresa de extensão rural (administração ou controle de processos) e a área de aplicação do software. Logo, para a análise de perfil, as possíveis combinações entre a empresa de extensão rural amostrada e cada conjunto de características observadas chegaram a 4.998.

Aplicando-se o método EM para verificar quantos perfis, teríamos, nessa mostra, o resultado de 24 perfis; e, a seguir, aplicando-se o método k-means, com distância de Manhattan, e conhecido o k (24 perfis), tem-se as modas de cada característica para cada perfil na Tabela 4.6. Como algumas características foram comuns aos

Tabela 4.6. Modas das características dos 24 perfis de empresas de extensão rural (2010).

Partição	UF	Problemas encontrados	Motivos investimento	Tipo área aplicação	Área aplicação
523 (10%)	SP	Dificuldade ou demora de resposta da Assistência Técnica quando foi solicitada	Capacitação dos usuários da instituição	Administração/gerenciamento	Comercialização
246 (5%)	SP	Linguagem/termos muito técnicos	Aquisição de equipamentos/hardware para disponibilizar aos produtores rurais	Administração/gerenciamento	Administração Rural
620 (12%)	SP	Falta de funcionários treinados	Capacitação dos usuários da instituição	Controle de processos	Irrigação
225 (5%)	SP	Dificuldade ou demora de resposta da Assistência Técnica quando foi solicitada	Aquisição de equipamentos/hardware para a própria instituição	Administração/gerenciamento	Contabilidade
205 (4%)	RO	Falta de funcionários treinados	Capacitação dos usuários da instituição	Administração/gerenciamento	Gerenciamento de pessoas
668 (13%)	SP	Subutilização das funcionalidades do software	Aquisição de equipamentos/hardware para a própria instituição	Controle de processos	Irrigação
106 (2%)	RO	Subutilização das funcionalidades do software	Capacitação dos usuários da instituição	Controle de processos	Irrigação
286 (6%)	SP	Falta de funcionários treinados	Aquisição de equipamentos/hardware para a própria instituição	Controle de processos	Irrigação
102 (2%)	RO	Linguagem/termos muito técnicos	Aquisição de equipamentos/hardware para a própria instituição	Controle de processos	Bioinformática
204 (4%)	SP	Complexidade do software	Desenvolvimento de software agropecuário próprio para utilização na instituição	Controle de processos	Previsão de safra
121 (2%)	SP	Subutilização das funcionalidades do software	Desenvolvimento de software agropecuário próprio para auxiliar no trabalho de Extensão rural, disponibilizando aos produtores rurais	Controle de processos	Previsão de safra
125 (3%)	SP	Complexidade do software	Aquisição de equipamentos / hardware para a própria instituição	Administração/gerenciamento	Base de dados.
164 (3%)	SP	Falta de funcionários treinados	Contratação de Assistência Técnica especializada para manutenção de hardware/software	Controle de processos	Previsão de safra
51 (1%)	BA	Outros	Desenvolvimento de software agropecuário próprio para utilização na instituição	Controle de processos	Administração Rural
161 (3%)	SP	Subutilização das funcionalidades do software	Aquisição de software agropecuário para auxiliar no trabalho de Extensão rural	Administração/gerenciamento	Administração Rural

Continua...

Tabela 4.6. Continuação...

Partição	UF	Problemas encontrados	Motivos investimento	Tipo área aplicação	Área aplicação
87 (2%)	SP	Falta de funcionários treinados	Aquisição de equipamentos /hardware para a própria instituição	Controle de processos	Fitossanidade
260 (5%)	RJ	Taxas extras associadas	Desenvolvimento de software agropecuário próprio para utilização na instituição	Controle de processos	Instrumentação agropecuária (instrumentos de medidas e amostragem)
211 (4%)	SP	Subutilização das funcionalidades do software	Aquisição de equipamentos/hardware para a própria instituição	Administração/gerenciamento	Outros
243 (5%)	SP	Falta de funcionários treinados	Aquisição de software agropecuário para auxiliar no trabalho de Extensão rural	Administração/gerenciamento	Contabilidade
13 (0%)	SP	Complexidade do software	Desenvolvimento de software agropecuário próprio para utilização na instituição	Controle de processos	Irrigação
119 (2%)	SP	Falta de funcionários treinados	Desenvolvimento de software agropecuário próprio para auxiliar no trabalho de Extensão rural, disponibilizando aos produtores rurais	Administração/gerenciamento	Gerenciamento de pessoas
11 (0%)	RO	Linguagem/termos muito técnicos	Aquisição de equipamentos/hardware para a própria instituição	Controle de processos	Previsão de safra
107 (2%)	RJ	Falta de funcionários treinados	Capacitação dos usuários da instituição	Controle de processos	Solos (análise química e física)
140 (3%)	SP	Subutilização das funcionalidades do software	Aquisição de software agropecuário para auxiliar no trabalho de Extensão rural	Controle de processos	GIS/GPS (geoprocessamento)

Fonte: Banco de dados de demandas em software agropecuário, do projeto SW Agro.

24 perfis, elas não estão apresentadas na tabela, que, novamente, foram: possuir computador, acesso à internet, utilizar software e pretender investir, pois as modas de todas elas em todos os perfis foram “sim”. Deve-se lembrar de que os valores de modas de cada característica refletem apenas o valor representativo do grupo, isto é, um grupo cuja moda da UF seja SP pode conter uma UF como RS ou AM; a moda reflete o valor predominante da característica no grupo, não necessariamente todos os valores do grupo.

Não foi possível cruzar dados de cultivo e manejo animal com as necessidades apontadas pelas empresas de extensão rural. Mesmo assim, os nichos de produção de sistemas agropecuários e capacitação de pessoal para utilizá-los, bem como capacitação de pessoal para instalação e manutenção de infraestrutura de software e hardware, ficaram bastante nítidos.

Novamente, os principais problemas encontrados com os sistemas de software que vêm sendo utilizados são subutilização das funcionalidades e problemas em torno deste, tais como: complexidade, falta de treinamento dos funcionários e linguagem inadequada; a questão de taxas extras é menos importante, pelo que pode ser observado nos perfis apresentados na Tabela 4.6.

Além disso, mais de 50% das necessidades encontram-se em sistemas de controle de processos, com forte presença da questão de irrigação e previsão de safras. Aparecem outras áreas, que se mostram mais específicas, como zoneamento, manejo integrado de pragas, análise do solo, instrumentação agropecuária e fitossanidade.

Por fim, estão presentes em todos os perfis encontrados, tanto a necessidade de novos produtos quanto a certeza de investimentos, quer sejam em capacitação de pessoal, aquisição de equipamentos, melhoria da infraestrutura, contratação de assistência técnica e/ou em compra de novos produtos de software.

Uma característica predominante em alguns perfis é o investimento no desenvolvimento de software agropecuário próprio para auxiliar o trabalho de extensão rural, tanto ser disponibilizado aos produtores rurais ou para uso na própria instituição. Isso ocorre em aproximadamente 25% dos perfis encontrados. Logo, é um indício que o mercado de software não tem coberto a demanda desse caso específico. Ou seja, as necessidades da extensão rural mostram-se bem particulares, exigindo soluções específicas e a intenção de tentar supri-las individualmente.

4.4.3 Perfis das empresas ofertantes de software agropecuário

Nessa seção, mostra-se como foram agrupadas as empresas que fornecem software agropecuário. Procurou-se verificar quais são as características observadas que melhor diferenciam seus perfis de atuação. Então, embora se tenham apenas 162 empresas amostradas, cada qual pode apresentar mais de um perfil, dependendo de como se combinem as características observadas. Assim, para que se pudesse encontrar os vários perfis, o conjunto final de características consideradas foi: a UF da empresa, sua classificação por porte (pequena, média, grande), a origem do capital, o tipo de aplicativo (produto de software) por ela representado, a finalidade do produto de software (administração/gerenciamento), a área de aplicabilidade em manejo animal, a área de aplicabilidade em cultivo vegetal, a área de aplicabilidade em controle de processos, a plataforma de desenvolvimento, o tipo de interface do produto, formas de comercialização, formas de licenciamento e as principais barreiras de entrada no mercado. Logo, para a análise de perfil, as possíveis combinações entre as empresas amostradas e os conjuntos de características observadas chegaram a 128.775.

Aplicando-se o método EM para verificar quantos perfis teríamos nessa mostra, o resultado foi 7 perfis; e, a seguir, aplicando-se o método k-means, com distância de Manhattan, e conhecido o k (7 perfis), temos as modas de cada característica para cada perfil nas Tabelas 4.7 a 4.14.

Embora algumas características tenham sido comuns aos sete perfis, todas estão apresentadas nas Tabelas 4.7 a 4.14, pois respondem a algumas questões levantadas na análise dos demandantes, cooperativas ou empresas de extensão rural. Novamente, deve-se lembrar de que os valores de modas de cada característica refletem apenas o valor representativo do grupo, isto é, um grupo cuja moda da Interface seja Web pode conter interfaces textuais ou gráficas stand alone; a moda reflete o valor predominante da característica no grupo, não necessariamente todos os valores do grupo.

Tabela 4.7. Perfis das empresas privadas desenvolvedoras de software agropecuário (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Classificação	Micro-empresa	Pequena empresa	Micro-empresa	Micro-empresa	Pequena empresa	Pequena Empresa	Média empresa
Origem do capital	Nacional privado	Nacional privado	Nacional privado	Nacional privado	Nacional privado	Nacional privado	Nacional privado
	Nacional privado	Nacional privado	Nacional privado	Nacional privado	Nacional privado	Nacional privado	Nacional privado
Tipo de aplicativo	Vertical	Vertical	Vertical	Vertical	Vertical	Vertical	Vertical
Abrangência	Municipal	Nacional	Municipal	Municipal	Municipal	Municipal	Nacional
Produtos/serviços	-	Sim, e apenas software para o agronegócio	-	-	-	-	Sim, mas não apenas com produtos e serviços para o agronegócio
Principal produto	-	O principal software foi desenvolvido pela própria empresa	-	-	O principal software foi desenvolvido pela própria empresa	-	O principal software foi desenvolvido pela própria empresa
Única distribuição?	sim	sim	sim	sim	sim	sim	não

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

Tabela 4.8. Perfis das empresas privadas desenvolvedoras de software agropecuário, segundo a área de aplicação na categoria Administração e Gerenciamento (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Administração rural, exceto ERP	não	não	sim	não	não	sim	não
Base de dados	não	não	não	não	não	não	não
Comercialização	não	não	não	não	não	sim	sim
Contabilidade	não	não	não	não	não	não	não
De insumos	não	não	não	sim	não	sim	não
De laboratório	não	não	não	não	não	não	não
De pessoas	não	não	não	não	não	não	não
De manutenção de maquinário/equipamentos	não	não	não	não	sim	não	não
Outros	não	sim	sim	sim	sim	sim	sim

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

Nas Tabelas 4.7 a 4.14, nota-se que a maioria dos produtos de software são verticais, isto é, aplicam-se a toda a cadeia produtiva, não apenas a segmentos específicos dela, e, também, que há um predomínio de capital próprio e nacional, bem como de micro e pequenas empresas.

Aproximadamente 20% delas, em dois perfis, representam produtos de software para administração rural, enquanto que em dois outros perfis, aproximadamente 25%, com gerenciamento

Tabela 4.9. Perfis das empresas privadas desenvolvedoras de software agropecuário, segundo a área de aplicação na categoria Manejo Animal (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Aves	não	não	não	não	não	não	não
Abelhas	não	não	não	não	não	não	não
Bovinos de corte	não	não	não	não	não	não	não
Bovinos de leite	não	não	não	não	não	não	não
Bubalinos	não	não	não	não	não	não	não
Caprinos	não	não	não	não	não	não	não
Equinos	não	não	não	não	não	não	não
Frutos do mar	não	não	não	não	não	não	não
Ovinos	não	não	não	não	não	não	não
Peixes	não	não	não	não	não	não	não
Suínos	não	não	não	não	não	não	não
Outros	não	sim	não	sim	sim	sim	não

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

Tabela 4.10. Perfis das empresas privadas desenvolvedoras de software agropecuário, segundo a área de aplicação na categoria Cultivo Vegetal (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Açúcar e álcool	não	não	não	sim	sim	não	não
Algodão	não	não	não	sim	não	não	não
Arroz	não	não	não	sim	não	não	não
Café	não	não	não	sim	sim	não	não
Dendê	não	não	não	sim	não	não	não
Eucalipto	não	não	não	sim	não	não	não
Feijão	não	não	não	sim	não	não	não
Frutas	não	não	não	sim	sim	não	não
Girassol	não	não	não	sim	não	não	não
Hortaliças	não	não	não	sim	não	não	não
Mamona	não	não	não	sim	não	não	não
Milho	não	não	não	sim	sim	não	não
Sistemas agroflorestais	não	não	não	sim	não	não	não
Soja	não	não	não	sim	sim	não	não
Trigo	não	não	não	sim	não	não	não
Floricultura	não	não	não	sim	não	não	não
Outros	não	sim	não	sim	sim	sim	não

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

de insumos e só 5%, representadas por um único perfil, com administração de maquinários e equipamentos.

No quesito de controle de processos, só se encontra produto de software voltado para adubação e calagem, os demais processos de interesse, que estão presentes na demanda, não se encontram representados nos perfis encontrados para os ofertantes. Esse resultado é um indicio de que o mercado de oferta de software agropecuário tem um nicho a explorar na área de controle de processos agrícolas.

A questão de principais barreiras de entrada no mercado de software, segundo os sete perfis encontrados, parece estar concentrada em especialização da mão-de-obra para as empresas que produzem seus softwares, necessidade de financiamento ou desconhecimento do mercado. Em comum, parece que a principal barreira é a distribuição de um único produto. Não se pode, no

Tabela 4.11. Perfis das empresas privadas desenvolvedoras de software agropecuário, segundo a área de aplicação na categoria Controle de Processo e/ou de Atividades Rurais (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Adubação, calagem e aplicação de calcário	não	não	não	sim	não	não	não
Agricultura de precisão	não	não	não	não	não	não	não
Agrimensura e topologia	não	não	não	não	não	não	não
Agrometeorologia	não	não	não	não	não	não	não
Automação / secagem / beneficiamento	não	não	não	não	não	não	não
Bioinformática	não	não	não	não	não	não	não
Fitossanidade	não	não	não	não	não	não	não
Genético	não	não	não	não	não	não	não
GIS/GPS	não	não	não	não	não	não	não
Geoprocessamento	não	não	não	não	não	não	não
Instrumentação	não	não	não	não	não	não	não
Inventário florestal	não	não	não	não	não	não	não
Irrigação	não	não	não	não	não	não	não
Manejo ambiental	não	não	não	não	não	não	não
Manejo integrado de pragas	não	não	não	não	não	não	não
Manejo florestal reflorestamento	não	não	não	não	não	não	não
Mecanização	não	não	não	não	não	não	não
Pecuária de precisão	não	não	não	não	não	não	não
Colheita, processamento e armazenagem	não	não	não	não	não	não	não
Previsão de safra	não	não	não	não	não	não	não
Rastreabilidade	não	não	sim	não	não	não	não
Receituário agrônomo	não	não	não	não	não	não	não
Receituário veterinário	não	não	não	não	não	não	não
Análise de solo	não	não	não	não	não	não	não
Zoneamento agrícola	não	não	não	não	não	não	não
Outros	não	sim	não	sim	sim	sim	não

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

entanto, ter alguma evidência de como seria a realidade se essas empresas tivessem um melhor conhecimento do mercado ou como elas pretendem expandir sua atuação, pois esses dados não foram refletidos pelos questionários aplicados à amostra.

Porém, em um único perfil em que o produto de software predominante não se enquadra em administrativo, há uma grande aplicação a cultivos vegetais, aparentemente o foco é adubação e calagem e envolve também o manejo animal. No outro perfil em que a situação é semelhante, o produto de software é voltado para a manutenção de equipamentos, e, é o único perfil em que a forma de licenciamento envolve transferência de tecnologia.

Tabela 4.12. Perfis das empresas privadas desenvolvedoras de software agropecuário, segundo plataforma, interface e formas de comercialização (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Plataforma							
Windows	não	sim	sim	sim	sim	sim	sim
Linux	não	não	não	não	sim	não	não
Outros	não	sim	não	sim	sim	sim	não
Interface							
Gráfica	sim	sim	sim	sim	sim	sim	sim
Web	não	não	não	não	sim	sim	sim
Textual	não	não	não	não	não	não	não
Formas de comercialização							
Via internet	não	sim	sim	não	sim	não	não
Direta	sim	sim	sim	sim	sim	sim	sim
Indireta com agregação de valores	não	não	não	não	sim	não	não
Indireta sem agregação de valores	não	não	não	não	sim	não	não
Sob encomenda	não	não	não	não	sim	não	sim
Outros	não	não	não	não	não	não	não

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

Tabela 4.13. Perfis das empresas privadas desenvolvedoras de software agropecuário, segundo as formas de licenciamento de seus produtos (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Gratuito com código fonte fechado	não	não	não	não	não	não	não
Gratuito com código fonte aberto	não	não	não	não	não	não	não
Pagamento único sem manutenção	não	não	sim	não	sim	não	sim
Aluguel, ASP ou SaaS	não	não	não	não	não	não	não
Uso com manutenção	sim	sim	sim	sim	sim	sim	sim
Transferência de tecnologia	não	não	não	não	sim	não	sim
Outros	não	não	não	não	não	não	não

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

Tabela 4.14. Perfis das empresas privadas desenvolvedoras de software agropecuário, segundo as barreiras competitivas (2010).

Grupos UF	105 (28%) SP	98 (26%) SP	31 (8%) MG	49 (13%) SP	20 (5%) SP	45 (12%) SP	26 (7%) SP
Economia de escala	não	não	não	não	não	não	não
Concorrência	não	não	não	não	não	não	não
Necessidade de financiamento	não	não	não	não	sim	não	não
Desconhecimento do mercado	não	não	não	não	sim	não	não
Mão-de-obra qualificada	não	sim	não	não	sim	não	sim
Outros	não	não	sim	não	não	não	não

Fonte: Banco de dados da oferta de software agropecuário, do projeto SW Agro.

4.5 Considerações finais

Existem poucos trabalhos na literatura relacionados à oferta e à demanda de software agropecuário, notadamente sobre os cenários de adoção de tecnologia da informação pelo setor rural. Esse tema vem despertando interesse porque a tecnologia da informação é um dos pilares para a modernização da agricultura, justificado por contribuir no incremento da produtividade e da produção agrícola.

Esse estudo não garante uma boa representatividade de ofertantes e de demandantes de software agropecuário do mercado nacional de software, devido às limitações da amostra coletada. Entende-se que mais esforço precisa ser envidado nesse sentido. Em particular, pretende-se expandir a amostra existente de demandantes e ofertantes e, para aplicar os questionários, antes eles serão reavaliados, de tal forma que as variáveis consideradas no estudo sejam representativas na análise do relacionamento entre a demanda e oferta de software agropecuário. A reformulação dos questionários deve considerar os seguintes objetivos básicos:

- 1) Fornecer o detalhamento das empresas demandantes de soluções em tecnologia da informação e comunicação (TIC) no universo agropecuário, incluindo suas necessidades na forma de produtos e serviços.
- 2) Com base nas necessidades dos demandantes, ativar as redes de serviços de TIC e articular ofertantes de produtos de software e prestadores de serviços de cada localidade.

Espera-se que essa investigação possa estimular a realização de estudos futuros com o propósito de identificar os nichos de mercado e as necessidades de realinhamento dos projetos das empresas desenvolvedoras de software rural para os próximos anos. Espera-se ainda que esse estudo possa contribuir de alguma forma para o estabelecimento de cenários economicamente viáveis sobre a adoção de tecnologia da informação pelo setor rural.

Convém ressaltar que a opção tecnológica não deve ser, exclusivamente, avaliada pelos seus benefícios econômicos, refletidos no aumento da produtividade e da produção. Os aspectos ambientais e sociais e a forma de difusão devem ser analisados conjuntamente, caso se pretenda investir em prol do desenvolvimento agrícola sustentável. Nesse contexto, o processo de informatização assume um papel muito importante, pois devido à heterogeneidade estrutural da agricultura nacional, as formas de condução desse processo podem reduzir ou ampliar ainda mais as diferenças sociais existentes.

4.6 Referências

- AMO, S. A. Técnicas de mineração de dados. In: JORNADAS DE ATUALIZAÇÃO EM INFORMÁTICA, 23., 2004, Salvador. **Anais...** Salvador: Universidade Federal da Bahia, 2004, v. 2, p. 195-236.
- ANDRADE, L. P. **Procedimento interativo de agrupamento de dados**. 2004. 193 p. Tese (Mestrado em Engenharia Civil) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- ARRAES, N. A. M.; LYRA FILHO, C. **A oferta de software agropecuário no Estado de São Paulo**. In: FEIRA E CONGRESSO DE INFORMÁTICA APLICADA À AGROPECUÁRIA E AGROINDÚSTRIA, 1995, Juiz de Fora, MG, 1995.
- BRANDÃO, G. E.; MEDEIROS, J. X. de. Programa de C&T para o desenvolvimento do agronegócio CNPq. In: CALDAS, R. de A.; PINHEIRO, L.E.L.; MEDEIROS, J.X. de; MIZUTA, K.; GAMA, G.B.M.N. da; CUNHA, P.R.D.L.; KUABARA, M.Y.; BLUMENSCHIN, A.. (Ed.). **Agronegócio brasileiro: ciência, tecnologia e competitividade**. Brasília, DF: CNPq, 1998. p. 11-25.
- CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. **CRISP-DM 1.0: step-by-step data mining guide**. Illinois: SPSS, 2000. 78 p.
- COSTA, C. N.; OLIVEIRA, S. R. de M. Tecnologia da informação para auxiliar o manejo de rebanhos e a tomada de decisões. **Revista dos Criadores**, São Paulo, v. 68, n. 821, p. 32-34, nov. 1998.
- COUTINHO, L.; FERRAZ, J.C. (Coord.). **Estudo da competitividade da indústria brasileira**. São Paulo: Papirus, 1994. 510 p.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, Blackwell, v. 39, n. 1, p. 1-38, 1977.
- HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**, 2nd ed. San Francisco: Morgan Kaufmann Publishers, 2006. p. 770.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**, New York, v. 31, n. 3, p. 264-323-97, Sep., 1999.
- KOTLER, P. **Administração de marketing: análise, planejamento, implementação e controle**. 5. ed. São Paulo: Atlas, 1998.
- _____. **Marketing para o século XXI: como criar, conquistar e dominar mercados**. São Paulo: Futura, 1999.
- KOTLER, P.; ARMSTRONG, G. **Princípios de marketing**. 7. ed. Rio de Janeiro: LTC Editora, 1999. 527 p.
- MACHADO FILHO, O. M. **Exploração e análise de agrupamento de dados**. 2002. 59 p. Tese (Mestrado em Engenharia Civil) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- MORAES, D. R. S. **Inteligência computacional na classificação litológica**. 2004. 75 p. Tese (Mestrado em Engenharia Civil) - Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- PUNTAR, S. G., **Métodos e visualização de agrupamento de dados**. 2003. 125 p. Tese (Mestrado em Engenharia Civil) - Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- REZENDE, S. O.; PUGLIESI, J. B.; MELANDA, E. A.; DE PAULA, M. F. Mineração de dados. In: REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. São Paulo: Manole, 2005. p. 307-336.

ROCHA, A., CHRISTENSEN, C. **Marketing**: teoria e prática no Brasil. 2. ed. São Paulo: Atlas, 1999. 288 p.

WEINSTEIN, A. **Segmentação de mercado**. São Paulo: Atlas, 1995. 314 p.

WITTEN, I. H.; FRANK, E. **Data mining**: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005. 525 p.